

Novel Data Analytics Meets Conventional Container Shipping: Predicting Delays by Comparing Various Machine Learning Algorithms

Adrian VIELLECHNER

WHU – Otto Beisheim School of Management
Adrian.Viellechner@whu.edu

Stefan SPINLER

WHU – Otto Beisheim School of Management
Stefan.Spinler@whu.edu

Abstract

Supply chain disruptions are expected to significantly increase over the next decades. In particular, delay of container vessels is likely to escalate due to rising congestion from continued growth of container shipping and higher frequency of extreme weather events. Predicting these delays could result in significant cost savings from optimizing operations. Both academic research and container shipping industry, however, lack analytical solutions to predict delay. To increase transparency on delay, we develop a prediction model based on 315 explanatory variables, 10 regression models, and 7 classification models. Using machine learning algorithms, we obtain best results for neural network and support vector machine with a prediction accuracy of 77 percent compared to only 59 percent of a naive baseline model. Various shipping players including sender, carrier, terminal operator, and receiver benefit from the easy-to-use prediction model to optimize operations such as buffers in schedules and the selection of ports and routes.

1. Introduction

Container shipping plays a crucial role in global cargo transportation and connects the entire supply chain from production to final customer [1]. Over the last three decades, the container shipping industry has been characterized by significant growth of annual volumes as well as vessel size [2, 3]. Congestion at ports and maritime chokepoints such as the Suez Canal, however, represents a major issue causing shipping delay already today [4]. The continuously growing industry is expected to further increase this problem. Besides congestion at ports and maritime chokepoints, bad weather also causes delay of container vessels [4]. Extreme weather events often lead to severe disruptions in supply chains, in particular affecting transportation [5]. Notably, the frequency of extreme weather events is strongly increasing worldwide [6, 7]. The ability

of predicting disruptions in supply chains caused by extreme weather events would massively enhance the resilience of supply chains [8]. To this end, novel methods from predictive analytics based on machine learning (ML) can help [8].

In general, disruptions in supply chains mainly resulting from extreme weather events and congestion can be diverse. Here, we focus on delay of container vessels as one type of supply chain disruptions. In particular, we aim at investigating whether delay of container vessels between Europe and Asia can be predicted by statistically analyzing historic delay using ML algorithms. Shipping players such as sender, carrier, terminal operator, and receiver could benefit from our predictions and thus reduce the risk of supply chain disruptions. Knowing the delays of vessels would help senders to select transport modes and ports, carriers to fine-tune schedules and choose alternative routes, terminal operators to adjust vessel handling sequences and optimize loading and discharging operations, and receivers to adapt hinterland logistics.

2. Literature review

2.1. Theoretical background

Previous studies investigate how the risk of supply chain disruptions can be decreased. In this literature review, we focus on supply chain risk management (SCRM). Overall, SCRM approaches can be classified by four main categories: disruption risk management (DRM), operational risk control (ORC), disaster and emergency management (DEM), and logistics service risk analysis (LSRA) [9]. Neither ORC nor LSRA consider unexpected supply chain disruptions, for instance, caused by extreme weather events, in contrast to DEM and DRM. DEM relates to post-event decision support which is not the focus in this research. DRM as pre-event decision support includes predicting delay of container vessels. Prediction models using fuzzy rule-based Bayesian network as a hybrid decision technique already exist [4]. A further study presents

a prediction model applying data mining and the ML algorithm random forest [10]. Moreover, qualitative delay estimates resulting from ML such as random forest are provided in a comparison analysis of two container terminals [11]. Neural network as another ML algorithm is implemented for predicting delay to forecast required human resources more accurately for covering daily port operations [12]. Similarly, more efficient allocation of human resources at ports is approached by a data mining research that suggests a classification and regression tree (CART) model [13]. All these studies regarding DRM, however, limit improvements to terminal operations as they consider the operational planning level. In contrast, we focus on the strategic planning level supporting shipping players such as sender, carrier, terminal operator, and receiver.

2.2. Motivation for research

The scientific relevance of this study is explained by three steps. First, there is a need for solutions with respect to big data and predictive analytics. Choi et al. [8] and Hazen et al. [14] confirm this scientific demand to optimize business processes and to predict system performance. Second, more solutions regarding risk management such as early-warning systems for logistics are required [9, 8]. And third, many studies emphasize the high importance of both topics at the interface. Among many others, Choi et al. [9] confirm the research gap at the interface of predictive analytics and risk management for supply chain management. In particular, Salleh et al. highlight the interest of both academia and shipping industry for solutions to predict delay of container vessels [15]. In summary, this and all aforementioned directions from other studies build the foundation for this research.

2.3. Review of prediction algorithms

To prepare for the model development in the following section, we now provide an overview of suitable prediction algorithms. To this end, we categorize them by using two dimensions: classification versus regression and statistical learning versus no statistical learning. On the one hand, we determine that classification and regression are most relevant for predicting delay compared to other ML algorithms such as anomaly detection or clustering. While the model outcome of a classification algorithm such as logistic regression is either yes or no, regression algorithms such as linear regression always provide a certain value. For instance, a classification algorithm predicts if it rains tomorrow or not. In contrast, a corresponding regression algorithm forecasts the amount of precipitation in liter

per square meter. On the other hand, we differentiate prediction algorithms by interpretability resulting in simple and complex [16]. We argue that interpretability depends on the characteristic whether the prediction algorithm is based on statistical learning. Using this categorization, we classify our considered prediction algorithms (Figure 1).

	Classification	Regression
Statistical learning	<ul style="list-style-type: none"> Neural network Random forest Support vector machine 	<ul style="list-style-type: none"> Neural network Random forest Support vector machine Lasso regression Ridge regression Elastic net regression
No statistical learning	<ul style="list-style-type: none"> (Multinomial) Logistic regression 	<ul style="list-style-type: none"> Linear regression

Figure 1. Categorization of prediction algorithms.

Regarding **learning-based classification**, we consider various predictive ML algorithms. Neural network (NN) is a powerful supervised ML algorithm, however, it cannot be easily interpreted as it learns from hidden knowledge in the data resulting in a blackbox [17]. Random forest (RF) is another supervised ML algorithm. Multiple decision tree-based RF is robust in predicting and thus prevents the problem of overfitting [16]. Support vector machine (SVM) is a supervised ML algorithm as well that uses a kernel function building variable subsets to reduce complexity [16]. Four different kernels can be used: polynomial, radial, linear, and sigmoid. NN, RF, and SVM are relevant for **learning-based regression** as well. Furthermore, lasso regression (LasR), ridge regression (RR), and elastic net regression (ENR) are ML algorithms that use penalization. They aim at lowering coefficients to prevent the problem of multicollinearity. Regarding **non-learning-based classification**, we apply (multinomial) logistic regression (LogR). LogR determines probabilities for observations to be assigned to two categories (here delay or no delay) [16]. For **non-learning-based regression**, we consider linear regression (LinR). LinR is a statistical prediction algorithm that defines a linear function to predict the independent variable [17].

Overall, this literature review section distinguishes SCRM approaches and introduces relevant DRM studies. Furthermore, it points out the need for work at the interface of data analytics and logistics and provides an overview of suitable prediction algorithms.

3. Methodology and model development

In this section, we describe our approach for building a model to predict delay of any container shipment within the following 365 days after prediction. Figure 2 provides an overview of our applied methodology.

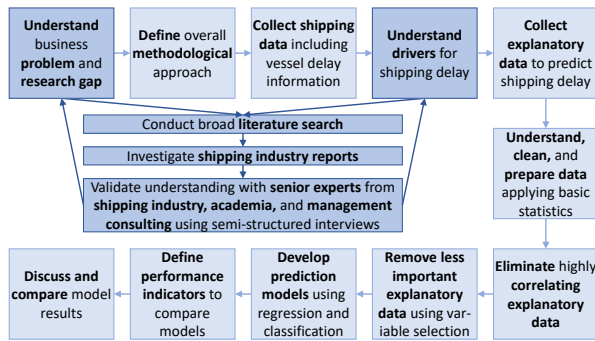


Figure 2. Overall methodology for this study.

To understand the business problem and to identify a research gap, we follow three steps. First, we conduct a broad literature search. Second, we obtain current industry-specific news from shipping reports. And third, we validate and challenge our findings from literature and industry news by performing several semi-structured interviews with highly experienced experts from shipping industry, academia, and top management consulting. After defining the overall methodological approach, we collect both shipping data including delay information and data that explains delay (3.1). Here, we apply aforementioned 3-step-approach again including literature search, industry news, and expert interviews to ensure understanding the drivers for delay. To reduce the number of collected explanatory variables, we conduct basic statistics such as correlation (3.2) and perform variable selection (3.3). Next, we develop prediction models that range from standard algorithms such as linear and logistic regression to more advanced ML algorithms such as random forest and neural network (3.4). Lastly, we present our selection of performance indicators (3.5) to evaluate models.

3.1. Data collection

To base the prediction of delay on a solid foundation, we access satellite data from the automatic identification system (AIS). AIS is non-public information used by ports and other maritime authorities to track vessels. There are online service providers such as MarineTraffic that offer paid services to access AIS data. Overall, we collect actual data from 75,814 container shipments departing from or arriving to Asian or European ports between February 2016 and August 2018. We only consider direct shipments without transshipment in between. To ensure including all long-distance shipments with transshipment and multiple port calls, we split them into several direct shipments. For instance, a shipment from Shanghai to Hamburg via Singapore is considered as two separate direct

shipments. In particular, we focus on 54,908 shipments that only connect European, Asian, and African ports. Most importantly, AIS data contains information on origin port (OP), destination port (DP), actual departure time, and actual arrival time.

To obtain the variable shipping delay, we compare actual information from AIS with scheduled shipping data. Here, we access the platform eeSea providing scheduled departure and arrival times from shipping services [18]. eeSea offers paid services with free trials to explore its profound data base. To ensure high data quality, we only consider shipping services connecting ports which are among the 100 busiest ports worldwide regarding total annual container throughput. After cleaning actual and scheduled data, we identify 2,954 shipments respectively observations with highest data quality for which we then calculate the variable shipping delay. It captures the absolute difference comparing actual and scheduled travel time of a vessel from its departure at origin port to its arrival at destination port. Besides actual and scheduled shipping data, we collect a large set of explanatory variables. To this end, we conduct a broad literature search followed by an investigation of current shipping reports primarily resulting in port congestion, ports inefficiencies, vessel issues, bad weather, and unreliability of the terminal operator [4]. To verify and challenge this set of variables, we interview six senior experts from leading carriers, academia, and top management consulting firms. According to the experts, the most important influencing factors for delay are bad weather at ports and en-route, port inefficiencies while loading or discharging the vessel, technical issues with the vessel, congestion at maritime chokepoints such as the Suez Canal or at ports, and forced waiting time at chokepoints or ports caused by tide conditions or missed slots. Additionally mentioned influencing factors are delay at previous ports, seasonality, piracy risk, and strikes. To account for the expert interviews, we enlarge our collection of explanatory variables. In total, we collect 315 explanatory variables which can be summarized by the following ten variable groups (Figure 3).



Figure 3. Illustration of explanatory variables.

Weather and natural disasters (146 variables): For weather, we consider precipitation and wind each with

maximum, minimum, and average values at OP, during travel, and at DP. We assess wind in more detail by analyzing ten different wind types. We match wind information with exact location and time of the vessel to understand if weather during travel influences delay [19]. For natural disasters at OP and DP, we include drought, earthquake, epidemic, extreme temperature, flood, landslide, storm/cyclone, volcanic, and wildfire and evaluate them by appearance risk, annual number of events, number of deaths, number of affected people, economic damage, and insured losses [20, 21]. **Demographics (64 variables):** For both OP and DP, we collect GDP growth, population growth, education level, labor productivity, strikes, unemployment rate, purchaser index, and 25 more [22, 23]. **Chokepoint congestion (37 variables):** Here, we investigate what and how many chokepoints each vessel passes during travel considering the European-Asian chokepoints Strait of Gibraltar, Danish straits, Bosphorus, Suez Canal, Strait of Bab El Mandeb, Strait of Hormuz, and Strait of Malacca. For particularly important Suez Canal and Strait of Malacca, we further analyze the number of vessels passing the chokepoint and 21 more. **Port congestion (20 variables):** For both OP and DP, we collect median time at port, median time at anchorage, number of vessels departing or arriving at port, and 12 more [24]. **Travel details (19 variables):** Here, we include delay at previous port, time at origin port, time between ports, and 17 more. **Piracy risk (10 variables):** For piracy risk, we evaluate number of attacks in high-risk areas and whether the vessel passes them. We identify five high-risk areas between Europe and Asia: Strait of Malacca together with Java Sea and Sulu Sea, South and East China Sea, Bay of Bengal with Arabian Sea, Red Sea with Gulf of Aden, and Persian Gulf with Gulf of Oman [25]. **Vessel characteristics (8 variables):** Regarding the vessel itself, we collect information on total size, age, flag, type, gross tonnage, deadweight, length, and breadth [26]. **Seasonality (5 variables):** We collect delay information from the same year, season, month, week, and calendar week. **Port size (4 variables):** For both OP and DP, we investigate absolute annual container volume and its relative annual growth [27]. **Shipper and service information (2 variables):** Here, we consider name of shipping company and name of shipping service [18].

3.2. Variable reduction using basic statistics

After completing data collection, we summarize all data in one master data sheet in Microsoft Excel which is then read into RStudio (Version 1.1.456). Note that we use R for all following statistical purposes. To ensure

a reliable data foundation for the model, we clean and preprocess data. To this end, we remove incomplete observations caused by variables with missing values leading to 1,851 remaining complete observations. Therefore, we exclude 53 explanatory variables with variance of 0 resulting from eliminating incomplete observations as these variables do not further contribute. Next, we investigate correlation. By using a threshold of 90 percent, we eliminate 65 strongly correlating variables. We exclude additional 26 variables by applying a threshold of 80 percent. Furthermore, we eliminate five factor variables such as name and country of ports as these are already covered by the total set of 64 demographics variables. This becomes clear when considering the general pattern of higher port inefficiency in developing countries. By performing all these data cleaning steps, we reduce the total number of explanatory variables from 315 to 166.

3.3. Variable reduction applying variable selection methods

We aim at further eliminating explanatory variables to reduce computation time, increase learning accuracy, and simplify the model without lowering the performance [28]. Note that we use the term variable selection for consistency reasons while other studies call it feature selection. In literature, three categories of variable selection methods are differentiated: filter, wrapper, and embedded methods [29]. Filter methods simply assign a score to each variable and hence are quick in computation but do not learn [28]. On the contrary, the slower but more advanced wrapper methods leverage learning as they test the performance of variable subsets to rank the individual variables [29]. Lastly, embedded methods determine variable importance by developing and training the model and thus are based on learning [28]. Overall, learning-based embedded methods are more capable than filter methods and more efficient than wrapper methods as they automatically evaluate variables while building the model [29]. Therefore, we apply embedded methods for variable selection and use ML-based lasso regression (LasR) for this purpose [30]. LasR with the method glmnet aims at lowering coefficients by penalizing variables [30]. The method glmnet is considered to be very efficient as it can handle large numbers of both observations and explanatory variables and thus is often chosen for variable selection [30].

To confirm validity of the variable selection according to LasR, we also investigate variable importance determined by random forest (RF) as another embedded

method [30]. As a result, RF substantially validates the variable selection derived from LasR. In detail, 75 percent of the 20 variables with highest importance according to RF are selected by LasR as well. Moreover, both LasR and RF assign highest variable importance to the same variable.

We apply LasR with the following custom control parameters: method = repeatedcv, number = 10, repeats = 5, and verboseIter = T. Further parameters for LasR include train/test-split = 0.7, method = glmnet, and standardize = FALSE. We run LasR twice by using two different data scaling methods to ensure that important variables are kept while less important variables are eliminated. Data can be either scaled by transforming the mean value to 0 and the standard deviation to 1 or by assigning all values to the range from 0 to 1. To run LasR, we only consider numerical variables as LasR would assign values to each level of factor variables. To account for these variables as well, we add them one-by-one and test the performance. After removing all variables with low importance, we run LasR once again to show the importance of the remaining numerical variables (Figure 4).

The ten most important variables according to LasR are the following. **Travel details:** Time between ports shows highest influence on shipping delay. Interestingly, it strongly correlates with number of chokepoints passed per vessel which we eliminate due to high correlation. Consequently, both travel time and number of chokepoints passed per vessel highly matter.

Piracy risk: Probability of passing Strait of Malacca, Java Sea, or Sulu Sea with many piracy attacks.

Demographics: Surprisingly, many demographics variables turn out to be highly important. This can be traced back to the fact that we consider demographics for both countries of OP and DP of a vessel journey. This implies that the current development state of a country influences efficiency of port operations and thus delay of container vessels. Most important ones are crops, basic education, deathrate, population density, and purchaser index. **Weather and natural disasters:**

High risk of cyclones exists in the Northwest Pacific Basin which is east of Vietnam and China. **Chokepoint congestion:** Most crucial chokepoint is the Strait of Bab El Mandeb that is south of Suez Canal.

To illustrate the most important explanatory variables for shipping delay, we introduce a brief shipping example of a container vessel. Assume it departs in Europe, transships in Indonesia as well as in the Philippines, and arrives in Hong Kong. This vessel faces high risk of delay as it travels very long distance and passes the Strait of Bab El Mandeb respectively Suez Canal where it might need to queue. It continues

through the Strait of Malacca followed by Java and Sulu Sea where significant piracy risk is present. It calls ports in Indonesia and the Philippines where it experiences less efficient port operations and thus higher loading and discharging times. On its remaining way to Hong Kong, it crosses the Northwest Pacific Basin confronting substantial risk of cyclones both en-route and for the arrival at Hong Kong port.

To provide more insights for the most important explanatory variables, we display relevant statistical information for them in Table 1.

Table 1. Statistical values for important variables.

Variable, unit	SD	5%ile	Median	95%ile
Time, days	4.5	0.5	1.9	15.4
Piracy M., 0/1	0.4	0.0	0.0	1.0
Crops DP, %	5.4	0.0	1.0	17.6
Bas. educ., %	8.7	12.5	22.4	42.3
Deathrate, %	0.2	0.4	0.9	1.1
Cyclone, 0/1	0.5	0.0	0.0	1.0
Crops OP, %	4.5	0.0	1.3	16.8
Str. Bab., 0/1	0.3	0.0	0.0	1.0
Pop. density, ppl. p. sq. mi.	1526	13.8	230.9	6482
Pur. in., 0-100	2.9	46.8	51.1	57.2

3.4. Model development

In the following, we present our developed prediction models by using the categorization as introduced in 2.3: classification versus regression and statistical learning versus no statistical learning. Regarding **learning-based classification**, we implement NN by applying train/test-split = 0.7, act.fct = logistic, 10 repetitions, 1 hidden layer with 1 neuron, and a threshold of 0.35 for the confidence matrix. We use R package neuralnet (version 1.44.2) with function neuralnet to develop NN [31]. For RF, we set train/test-split = 0.7, ntree = 1100, mtry = 7, importance = false, and proximity = true. We use R package randomForest (version 4.6-14) with function randomForest to develop RF [30]. For SVM, we explore all four kernels polynomial, radial, linear, and sigmoid of which polynomial results in highest performance. We apply train/test-split = 0.7, cost = 1000, gamma = 0.01, coef0 = 1, degree = 2, and epsilon = 0.1. We use R package e1071 (version 1.7-1) with function svm to develop SVM [31]. Regarding **learning-based regression**, we implement NN, RF, and SVM as well. For NN, we use train/test-split = 0.8, 10 repetitions, and 1 hidden layer with 4 neurons. For RF, we set the following parameters: train/test-split = 0.8, ntree

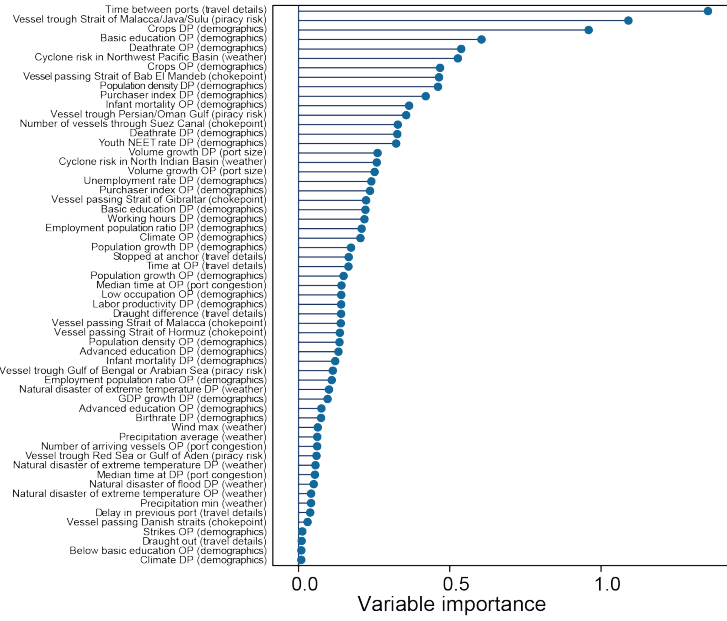


Figure 4. Variable importance derived from lasso regression.

= 1000, mtry = 59, and nodesize = 5. For SVM, we apply train/test-split = 0.7, cost = 10, gamma = 0.01, coef0 = 1.5, degree = 3, and epsilon = 0.01. Furthermore, we run LasR, RR, and ENR for which we configure train/test-split = 0.7 and customer control parameters such as number = 10 and repeats = 5. We use R package glmnet (version 2.0-16) with function glmnet to develop LasR, RR, and ENR [30]. Regarding **non-learning-based classification**, we implement LogR for which we apply a train/test-split of 0.7 and a threshold of 0.4 for the confidence matrix. We use R package stats (version 3.5.1) with function glm to develop LogR [32]. Regarding **non-learning-based regression**, we run LinR for which we set train/test-split to 0.7 and apply the step function to identify a smaller variable subset. We use R package stats (version 3.5.1) with function lm to develop LinR [32].

3.5. Performance indicators

To evaluate regression models, we primarily use Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Both help comparing predicted with actual values [33] and are calculated as follows [34].

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (\text{predicted}_n - \text{actual}_n)^2}{N}} \quad (1)$$

$$MAE = \frac{\sum_{n=1}^N |\text{predicted}_n - \text{actual}_n|}{N} \quad (2)$$

To evaluate classification models, we primarily apply RMSE, accuracy, and sensitivity. We explain the latter two by illustrating a confidence matrix (Figure 5).

		Predicted	
		Negative	Positive
Actual	Negative	True negative (TN)	False positive (FP)
	Positive	False negative (FN)	True positive (TP)

Figure 5. Confidence matrix for classification models.

In shipping, positive relates to delay while negative means on-time. Thus, accuracy shows the correctness of both positive and negative delay predictions relative to all delay predictions (Equation 3).

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (3)$$

Sensitivity particularly focuses on positive actual values. In the shipping context, sensitivity explains how often we predict delay in case the vessel arrives with delay (Equation 4). Note that sensitivity is often also named recall or true positive rate in literature.

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

With respect to container shipping, we argue that it is

more expensive to falsely predict on-time if the vessel arrives delayed than to falsely predict delay if the vessel arrives on-time. In other words, an unexpected delayed arrival is more difficult to handle than an unexpected on-time arrival. For instance, falsely predicting delay can be corrected by contacting the captain to confirm the delay. On the contrary, falsely predicting on-time certainly results in issues including the adaption of the vessel handling sequence and hinterland logistics. In summary, this section on methodology and model development explains the overall methodological approach, describes all data collection steps, and elaborates on variable reduction. Ultimately, it shows the development of all models and prepares discussion by introducing relevant performance indicators.

4. Model results

In this section, we describe and compare all developed regression (4.1) and classification models (4.2) to select the best prediction model.

4.1. Discussion of regression models

We can now evaluate and compare performance of developed regression models as summarized in Table 2.

Table 2. Results of regression prediction models (with performance values and column ranks).

Model	RMSE	MAE
SVM polynomial	0.43 (1)	0.26 (1)
NN	0.52 (2)	0.34 (2)
SVM radial	0.53 (3)	0.34 (2)
RF	0.63 (4)	0.40 (4)
LinR	0.67 (5)	0.49 (6)
SVM linear	0.70 (6)	0.47 (5)
LasR	0.79 (7)	0.56 (7)
ENR	0.79 (7)	0.56 (7)
RR	0.80 (9)	0.56 (7)
SVM sigmoid	0.93 (10)	0.61 (10)

According to the selected performance indicators, we argue that SVM with polynomial kernel clearly achieves highest prediction performance, followed by NN with four neurons in one hidden layer and SVM with radial kernel. Interestingly, all three statistical learning-based models perform significantly better regarding both RMSE and MAE than non-learning-based LinR and other learning-based models. While we can easily compare the performance of the models relatively, it is more difficult to evaluate the performance of the models in general due to the low interpretability of performance

indicators. However, Figure 6A reveals that NN generally performs well when comparing predicted and actual delay whereas Figure 6B shows that SVM with polynomial kernel predicts even better.

4.2. Discussion of classification models

Besides regression models, we can now also evaluate and compare performance of developed classification models as summarized in Table 3.

Table 3. Results of classification prediction models (with performance values and column ranks).

Model	RMSE	Accuracy	Sensitivity
NN	0.41 (1)	0.77 (3)	0.78 (2)
LogR	0.41 (1)	0.75 (5)	0.75 (3)
RF	0.43 (3)	0.81 (1)	0.65 (5)
SVM polynom.	0.46 (4)	0.79 (2)	0.68 (4)
SVM radial	0.48 (5)	0.77 (3)	0.64 (6)
SVM linear	0.50 (6)	0.74 (6)	0.48 (7)
SVM sigmoid	0.61 (7)	0.63 (7)	0.96 (1)

According to the table, NN and LogR achieve best (lowest) value for RMSE whereas RF scores best (highest) for accuracy. While NN and LogR are equal in RMSE, NN outperforms LogR in accuracy and sensitivity, thus we prefer NN. To identify the best prediction model comparing NN with RF, we need to further investigate their results. Both models perform similarly well regarding RMSE, but strongly differ with respect to accuracy and sensitivity. On the one hand, NN misclassifies delay versus no delay 127 times resulting from 81 false positives (FP) and 46 false negatives (FN) (Figure 7A). We compare this to RF with only 108 misclassifications resulting from 36 FP and 72 FN (Figure 7B). For this reason, the accuracy for RF with 81 percent is higher than for NN with 77 percent. Notably, the accuracy of the baseline model is only 59 percent meaning one could simply always predict no delay resulting in an accuracy of 59 percent. Thus, both RF and NN achieve significantly higher accuracy than the baseline model. On the other hand, NN results in only 46 FP compared to RF with 72 FP leading to higher sensitivity for NN. As already highlighted in 4.1, we argue that it is more important to prevent unexpected delayed arrivals of vessels than to prevent unexpected on-time arrivals. Therefore, we prefer NN. Overall, we claim that NN performs best regarding classification models, followed by LogR and RF due to the particular importance of sensitivity. Surprisingly, non-learning-based LogR performs almost as well as statistical learning-based NN, the selected best classification model.

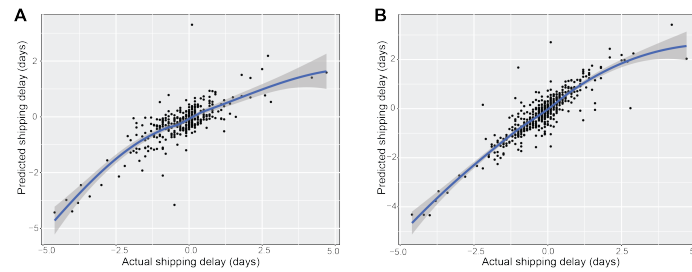


Figure 6. Comparison of actual and predicted delay for NN (A) and SVM (B) as best regression models.

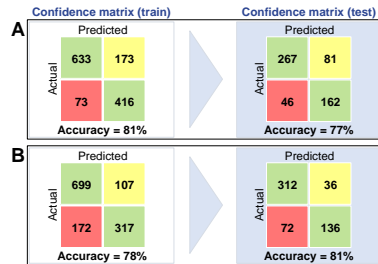


Figure 7. Confidence matrix for NN (A) and RF (B) as best classification models.

Interestingly, the evaluation of classification models helps to better understand the performance of developed regression models as we always apply the performance indicator RMSE. To this end, we compare RMSE for NN as best classification model to SVM with polynomial kernel as best regression model. As a result, we only obtain a small difference: 0.41 for NN versus 0.43 for SVM. Surprisingly, the regression model is almost as good as the classification model even though it attempts to predict delay precisely in contrast to all classification models which only classify delay versus no delay. In addition, we consider the aspect of interpretability when comparing NN and SVM to select the best model. Complex models such as NN can achieve high accuracy on the one hand but result in low interpretability on the other hand [16]. In summary, we argue that SVM with polynomial kernel serves as best prediction model in this study considering both performance and interpretability.

Overall, this model results section evaluates all developed models by first analyzing classification and regression separately. To select the best overall prediction model, best-performing classification and regression models are compared with each other.

5. Managerial implications

In general, we aim at emphasizing that the model supports the strategic planning level to draw long-term conclusions regarding operations. For instance, carriers

might want to increase buffers in their schedules for certain months in the year which regularly show more vessel delay caused by bad weather. On the contrary, the model does not focus on the operational planning level to make short-term adjustments such as reacting to bad weather forecasts for the next days.

Four different types of companies can benefit from the proposed prediction model. First, senders of goods can leverage more intermodal transportation for certain routes at specific times in the year to avoid congestion at maritime chokepoints. Intermodal transportation combines different means of transportation such as sea and rail freight and becomes more important particularly for transport between Europe and Asia [35]. In addition, senders can select other ports with less congestion and more efficient operations either temporarily for certain times of the year or permanently if the prediction model confirms less delay for nearby ports. Second, carriers which transport the goods can operate on alternative routes from Europe to Asia such as the Northern Sea Route going around Russia which has been gaining importance in recent years to avoid bottlenecks at Suez Canal and Strait of Bab El Mandeb [36]. Moreover, carriers can adjust buffers in their schedules for routes with high delay prediction. Third, terminal operators can optimize their loading and unloading operations including the sequence of vessels for handling. And forth, receivers of goods can reduce risks such as stock-outs in stores, insufficient supply for production, high capital cost, or lost sales depending on the business of the receiver.

To enable these adjustments of shipping operations, the aforementioned companies can extract different types of information from the prediction model. On the one hand, they benefit from predicted delay for their relevant shipping routes. Predicted delay is the final model outcome which helps them to adjust shipments with high risk of delay. Here, we emphasize that companies can benefit significantly more from SVM with polynomial kernel as best regression model compared to NN as best classification model. This becomes clear when recalling the granularity of the

predictions. Companies would be limited in their decision taking when only obtaining the information of vessels being delayed or on-time without more accurate estimations. For instance, receivers would still struggle to precisely optimize hinterland logistics and thus include substantial buffers in their operations.

On the other hand, companies benefit from intermediate model outcome such as variable importance. For instance, the variable time between ports is highly important and thus companies should include more buffer into schedules and following logistics regarding long-distance shipments. Moreover, a large number of demographics variables with high importance reveals that ports in developing countries often struggle with efficient operations. This also accounts for ports with high volume growth indicated by the port size variables. Consequently, fast-growing ports particularly in developing countries could be reconsidered. Next, important variables with fluctuation during the year such as cyclone risk in the Northwest Pacific Basin and chokepoint congestion in the Strait of Bab El Mandeb and Suez Canal can be investigated more closely. To this end, we recommend to understand the level of risk at different times in the year. Furthermore, many travel details variables directly affect delay on the same route. For instance, monitoring the variables time at origin port and delay at previous port allows for short-term adjustments of operations to react to emerging delay.

In general, the management of aforementioned companies can easily implement and use the proposed prediction model as it is entirely built in R, up-to-date, and ready-to-use. To effectively encounter delay in the future, companies are advised to follow three steps. First, companies should decide what prediction error type is more crucial for their business as prediction algorithms face different suitability with regards to prediction error types. For instance, if false negatives, meaning on-time predictions in case of delayed arrivals, are more expensive for the company, it is essential to select an algorithm with high sensitivity. Thus, the risk of obtaining false negatives is reduced. Second, the model should be fed with new shipments after a few years. This step is important as present circumstances might change in the future. For instance, shipments to or from Hamburg might face less delay in the future considering current expansion work at the river that serves Hamburg. It is expected that this expansion reduces port entry restrictions due to tide conditions for large vessels. And third, it is recommended that companies update a few explanatory variables as well. Here, the vast majority of variables remains effective, however, some variables such as the operating carrier need to be updated as carriers might improve or fall off

in on-time service quality over time.

Interestingly, the prediction model can be generally applied for other means of transportation such as rail transportation. Delay in rail transportation, however, is partly influenced by other factors. To this end, the selection of explanatory variables must be adjusted.

6. Conclusion and future directions

In this study, we provide data analytics-based solutions for the container shipping industry. We introduce a developed prediction model to increase transparency on delay of container vessels between Europe and Asia. NN as our best classification model achieves RMSE of 0.41 and prediction accuracy of 77 percent compared to only 59 percent in case of naive forecasting. SVM with polynomial kernel as our best regression model results in RMSE of 0.43. Notably, the RMSE of the regression model SVM which precisely predicts delay is higher by only 0.02 compared to the classification model NN which only classifies the vessels as delayed or on-time.

In summary, our study shows that various shipping players benefit from increased transparency on shipping delay such as sender, carrier, terminal operator, and receiver. As highlighted in the previous section, our model targets the strategic planning level to draw long-term conclusions regarding operations. This provides additional directions for future studies. A prediction model focused on short-term adjustments on a daily basis should include more real-time data such as weather forecasts and live congestion reports. That type of prediction model with more short-term benefits could be used in addition to the proposed model in this study within a corporate decision support system for the container shipping industry.

References

- [1] C.-Y. Lee, H. L. Lee, and J. Zhang, "The impact of slow ocean steaming on delivery reliability and fuel consumption," *Transportation Research Part E: Logistics and Transportation Review*, vol. 76, pp. 176–190, 2015.
- [2] N. K. Tran and H.-D. Haasis, "An empirical study of fleet expansion and growth of ship size in container liner shipping," *International Journal of Production Economics*, vol. 159, pp. 241–253, 2015.
- [3] M. A. Dulebenets, "A comprehensive multi-objective optimization model for the vessel scheduling problem in liner shipping," *International Journal of Production Economics*, vol. 196, pp. 293–318, 2018.
- [4] N. H. M. Salleh, R. Riahi, Z. Yang, and J. Wang, "Predicting a containership's arrival punctuality in liner operations by using a fuzzy rule-based bayesian network (FRBBN)," *The Asian Journal of Shipping and Logistics*, vol. 33, no. 2, pp. 95–104, 2017.

- [5] C. Speier, J. M. Whipple, D. J. Closs, and M. D. Voss, "Global supply chain design considerations: Mitigating product safety and security risks," *Journal of Operations Management*, vol. 29, no. 7-8, pp. 721–736, 2011.
- [6] B. Mahapatra, M. Walia, and N. Saggurti, "Extreme weather events induced deaths in India 2001–2014: Trends and differentials by region, sex and age group," *Weather and Climate Extremes*, vol. 21, pp. 110–116, 2018.
- [7] S. A. Markolf, C. Hoehne, A. Fraser, M. V. Chester, and B. S. Underwood, "Transportation resilience to climate change and extreme weather events – Beyond risk and robustness," *Transport Policy*, 2018.
- [8] T.-M. Choi, H. K. Chan, and X. Yue, "Recent development in big data analytics for business operations and risk management," *IEEE transactions on cybernetics*, vol. 47, no. 1, pp. 81–92, 2017.
- [9] T.-M. Choi, C.-H. Chiu, and H.-K. Chan, "Risk management of logistics systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 90, pp. 1–6, 2016.
- [10] J. Yu, G. Tang, X. Song, X. Yu, Y. Qi, D. Li, and Y. Zhang, "Ship arrival prediction and its value on daily container terminal operation," *Ocean Engineering*, vol. 157, pp. 73–86, 2018.
- [11] C. Pani, T. Vanelander, G. Fancello, and M. Cannas, "Prediction of late/ early arrivals in container terminals - A qualitative approach," *European Journal of Transport and Infrastructure Research (EJTIR)*, vol. 15, no. 4, pp. 536–550, 2015.
- [12] G. Fancello, C. Pani, M. Pisano, P. Serra, P. Zuddas, and P. Fadda, "Prediction of arrival times and human resources allocation for container terminal," *Maritime Economics & Logistics*, vol. 13, no. 2, pp. 142–173, 2011.
- [13] C. Pani, P. Fadda, G. Fancello, L. Frigau, and F. Mola, "A data mining approach to forecast late arrivals in a transshipment container terminal," *TRANSPORT*, vol. 29, no. 2, pp. 175–184, 2014.
- [14] B. T. Hazen, J. B. Skipper, C. A. Boone, and R. R. Hill, "Back in business: operations research in support of big data analytics for operations and supply chain management," *Annals of Operations Research*, vol. 270, no. 1-2, pp. 201–211, 2018.
- [15] N. H. M. Salleh, R. Riahi, Z. Yang, and J. Wang, "Supply chain risk management in the container liner shipping industry from a strategic point of view," *European Journal of Business and Management*, vol. 7, no. 24, pp. 155–163, 2015.
- [16] M. Saberioon, P. Císař, L. Labbé, P. Souček, P. Pelissier, and T. Kerneis, "Comparative performance analysis of support vector machine, random forest, logistic regression and k-nearest neighbours in rainbow trout (*oncorhynchus mykiss*) classification using image-based features," *Sensors (Basel, Switzerland)*, vol. 18, no. 4, 2018.
- [17] W.-J. Niu, Z.-K. Feng, B.-F. Feng, Y.-W. Min, C.-T. Cheng, and J.-Z. Zhou, "Comparison of multiple linear regression, artificial neural network, extreme learning machine, and support vector machine in deriving operation rule of hydropower reservoir," *Water*, vol. 11, no. 1, p. 88, 2019.
- [18] eeSea, "Liner services." URL: <https://www.eesea.com/modules/liner-services/> [accessed on January 16, 2019], 2018.
- [19] National Aeronautics and Space Administration, "Earthdata search." URL: <https://disc.gsfc.nasa.gov/> [accessed on February 04, 2019], 2018.
- [20] Centre for Research on the Epidemiology of Disasters - CRED, "EM-DAT: The international disaster database." URL: <https://www.emdat.be/database> [accessed on October 23, 2018], 2018.
- [21] US Department of Commerce, "JetStream: Introduction to tropical cyclones." URL: <https://www.weather.gov/jetstream/tc> [accessed on March 11, 2019], 2019.
- [22] G. Shackman, "Global social change reports." URL: <http://gsociology.icaap.org/dataupload.html> [accessed on October 16, 2018], 2015.
- [23] International Labour Organization, "ILOSTAT." URL: <https://www.ilo.org/ilostat/> [accessed on January 30, 2019], 2018.
- [24] MarineTraffic, "Port congestion." URL: <https://www.marinetraffic.com/en/online-services/single-services/port-congestion> [accessed on November 12, 2018], 2018.
- [25] ICC International Maritime Bureau, "Piracy and armed robbery against ships."
- [26] MarineTraffic, "Vessels database." URL: https://www.marinetraffic.com/en/data/?asset_type=vessels [accessed on January 30, 2019], 2019.
- [27] World Shipping Council, "Top 50 world container ports." URL: <http://www.worldshipping.org/about-the-industry/global-trade/top-50-world-container-ports> [accessed on June 06, 2019], 2017.
- [28] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, "Feature selection: An ever evolving frontier in data mining," *JMLR: Workshop and Conference Proceedings 10*, pp. 4–13, 2010.
- [29] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [30] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*. New York, NY: Springer New York, second edition ed., 2009.
- [31] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning: With applications in R*, vol. 103 of *Springer texts in statistics*. New York: Springer, 2013.
- [32] A. P. Field, J. Miles, and Z. Field, *Discovering statistics using R*. London and Thousand Oaks Calif.: Sage, 2012.
- [33] W. Wang and Y. Lu, "Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model," *IOP Conference Series: Materials Science and Engineering*, vol. 324, p. 012049, 2018.
- [34] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014.
- [35] H. Rodemann and S. Templar, "The enablers and inhibitors of intermodal rail freight between Asia and Europe," *Journal of Rail Transport Planning & Management*, vol. 4, no. 3, pp. 70–86, 2014.
- [36] D.-Y. Lin and Y.-T. Chang, "Ship routing and freight assignment problem for liner shipping: Application to the Northern Sea Route planning problem," *Transportation Research Part E: Logistics and Transportation Review*, vol. 110, pp. 47–70, 2018.